

Weekly Report

Yuxin Ma

06.25.2012 - 07.01.2012

Readings

Due to the graduation affairs, plans of some papers during May and June were shelved. This week I resumed reading these papers, mainly about the works from Stanford on data wrangling and graph visualization:

- *Research directions in data wrangling: Visualizations and transformations for usable and credible data;*
- *Graph Prism: Compact Visualization of Network Structure;*
- *Profiler: Integrated Statistical Analysis and Visualization for Data Quality Assessment*

Also the upcoming paper *Model-Driven Design for the Visual Analysis of Heterogeneous Data* has described a great structure of visual analysis flow and design.

Practice & Skills

I tried some demos from the official OpenMPI tutorial this week including the Hello World program of identifying process ID and core ID, and simple message passing functions between MPI processes. I glanced through some other tutorials as well on configurations of master-slave MPI clusters, which might be useful if multiple hosts are available to use.

Research

The edges has been retrieved from the two location-based datasets and transformed into GraphViz DOT format with $\sim 1,000,000$ edges and $\sim 100,000$ edges. The current issue focuses on how to find the mis-clustered node.

For the performance of Gephi, it has now been a problem that the demo system suffers a little bit time-tolerance for brushing and other interactions. Usually for the smaller Brightkite dataset the refresh rate of ParaView will be more than 1-2 minutes.

The raw result of Brightkite dataset is in the attachment.

Miscellaneous

None.

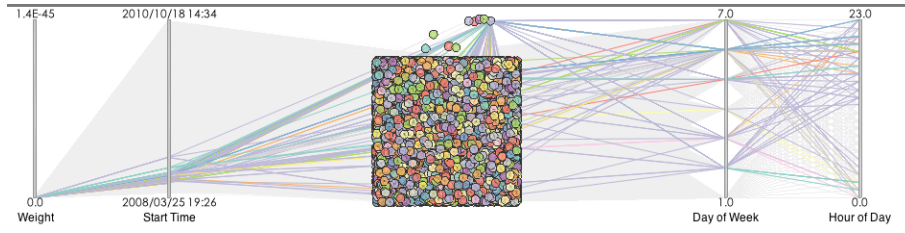


Figure 1: Brightkite dataset with $\sim 100,000$ separated edges and $\sim 10,000$ nodes. The current effect is highlighting paraedges of several node.

Plan for Next Week

- Help Dr. Dichao Peng with the paper;
- start plans of reviewing top researchers' foundational papers.